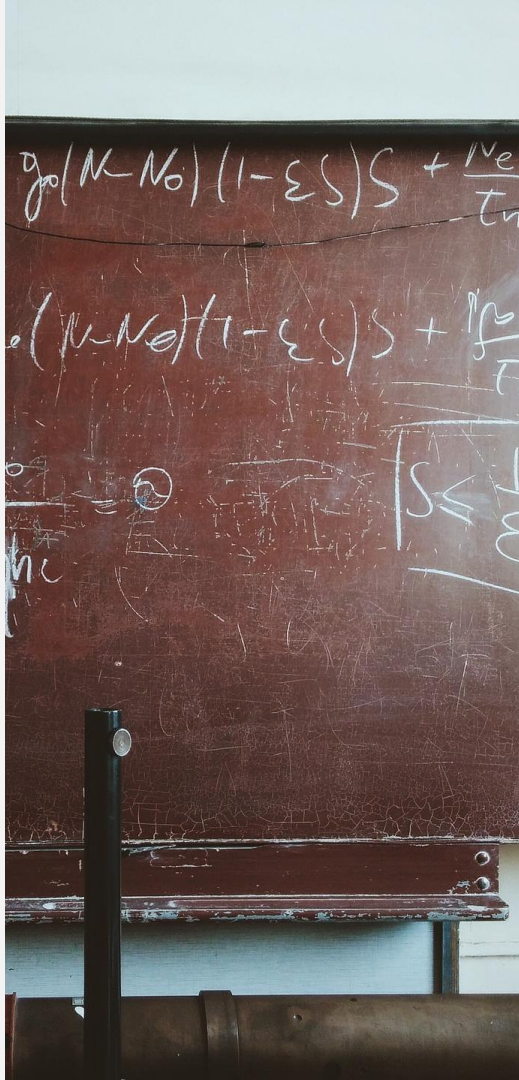
Análise Preditiva II

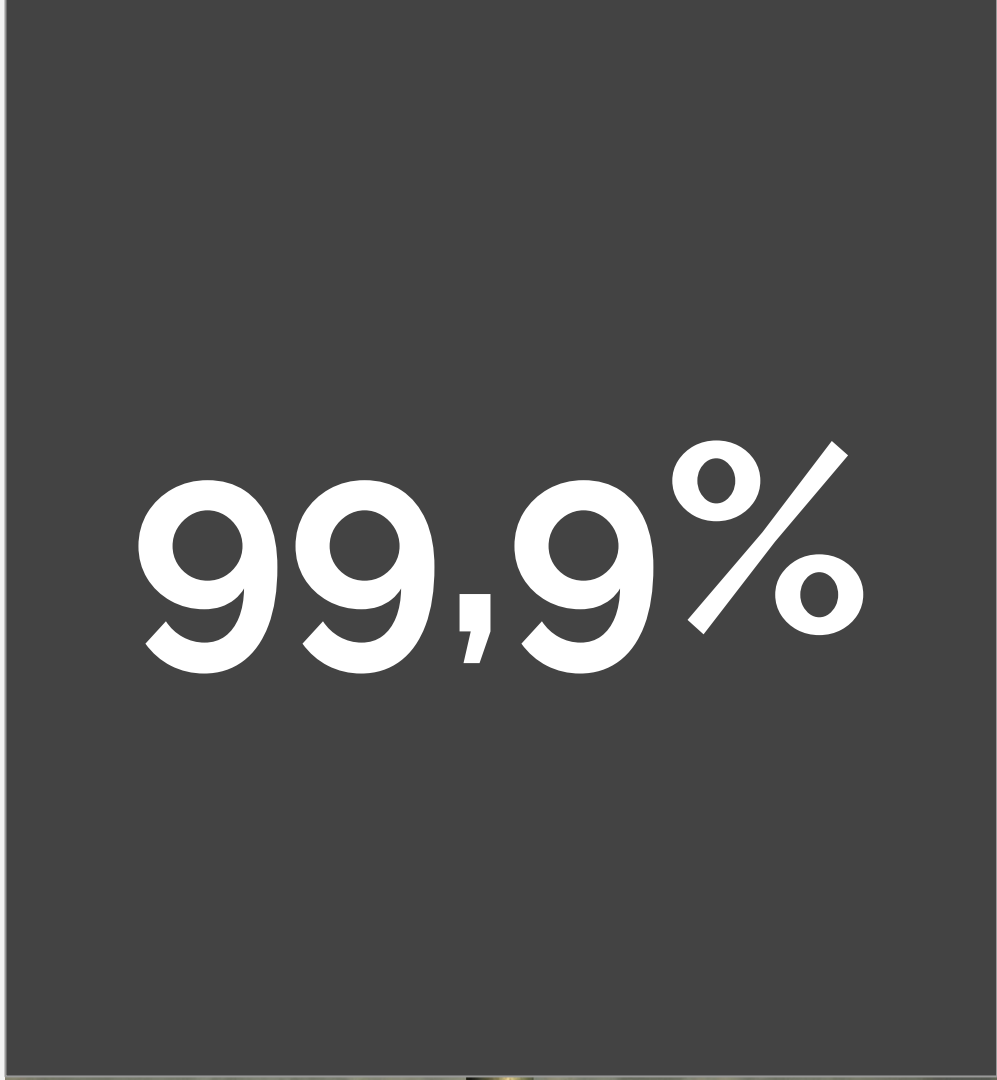
01

Overfitting e Underfitting



Overfitting e Underfitting

Você confiaria no seu modelo?

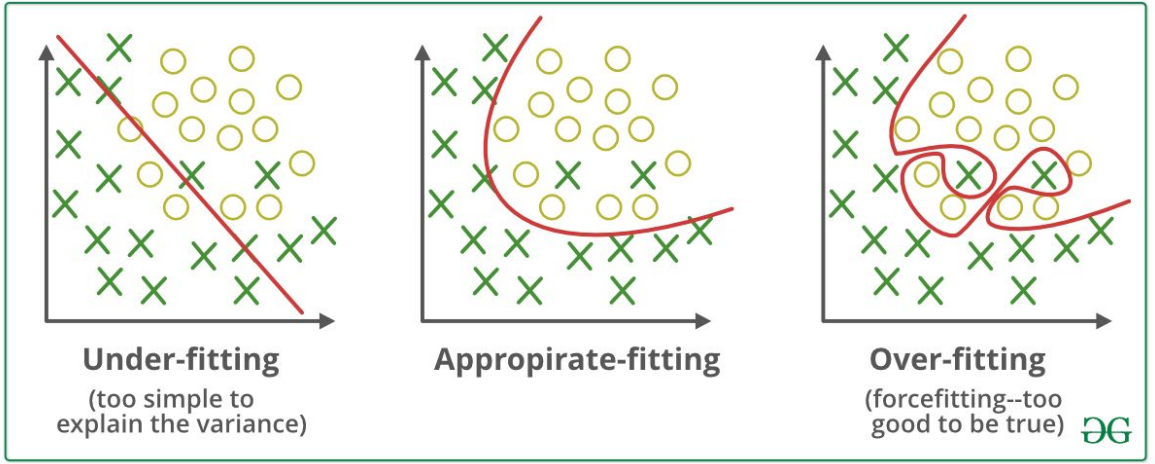


Underfitting

- Treino ruim
- Não encontra padrões

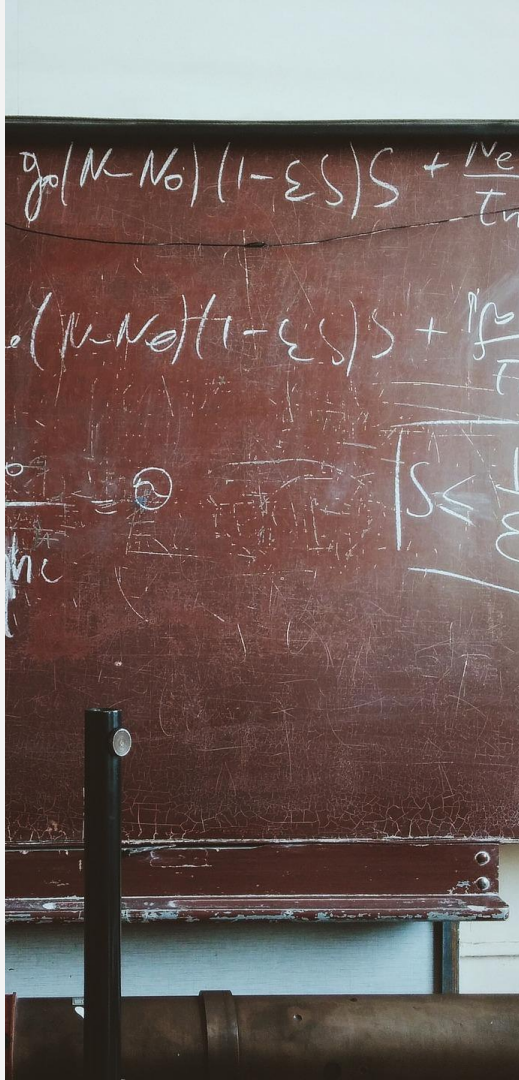
Overfitting

- Treino excelente
- Teste resultado ruim
- Problema em generalizar

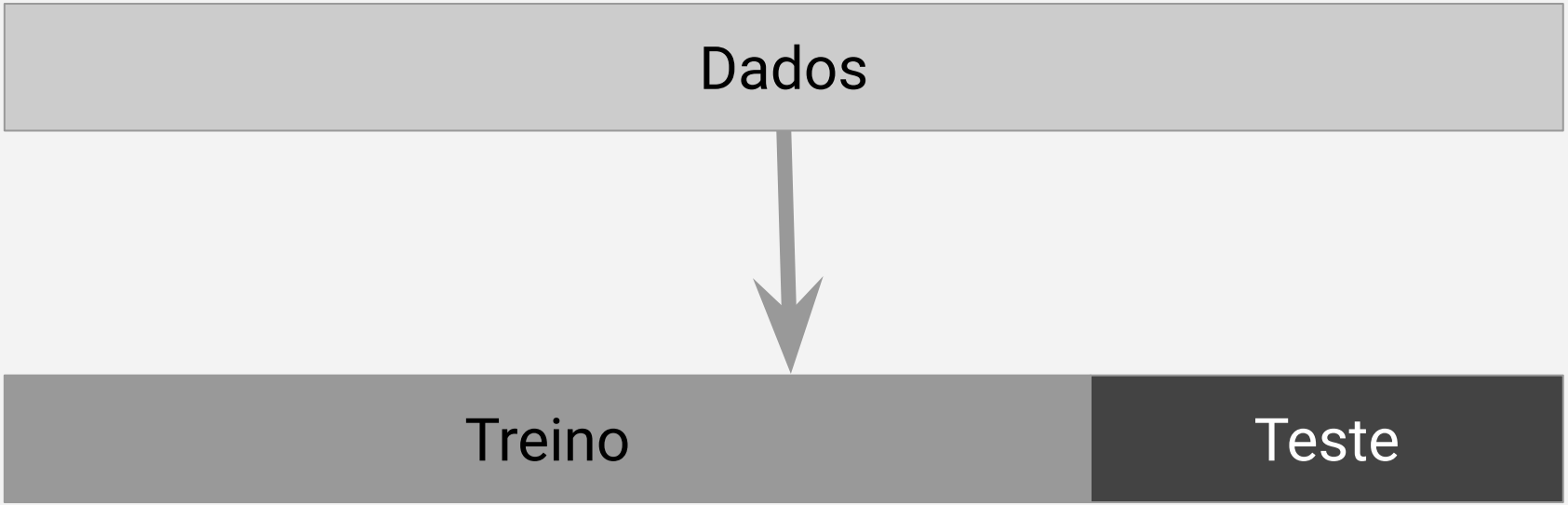


02

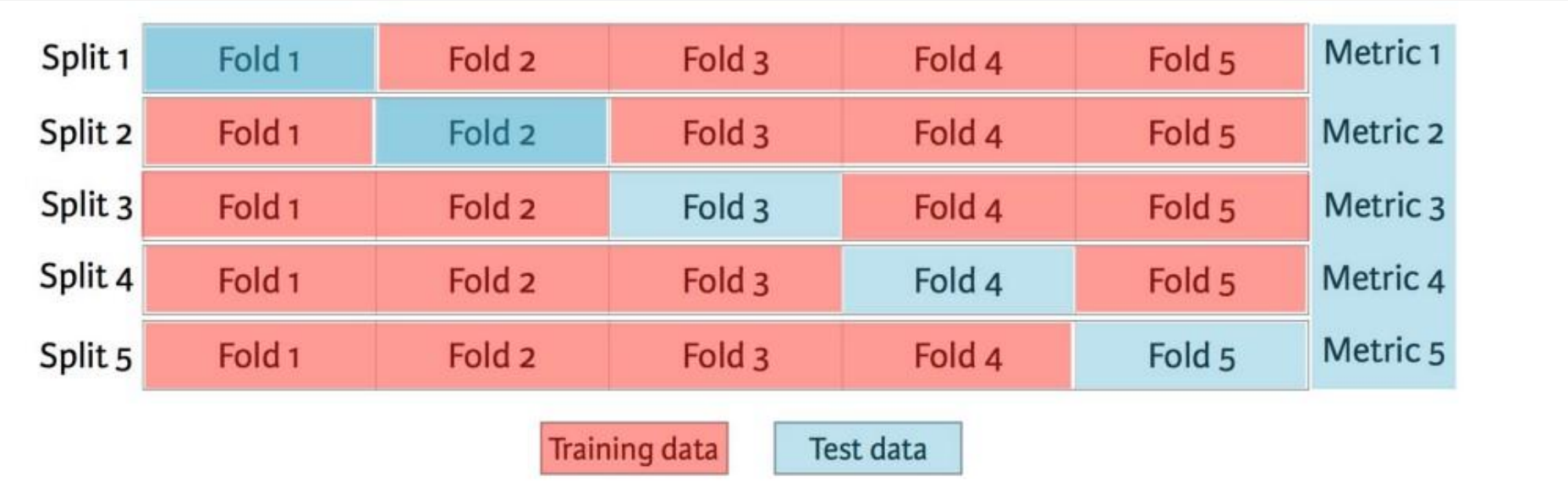
Validação



Hold-out

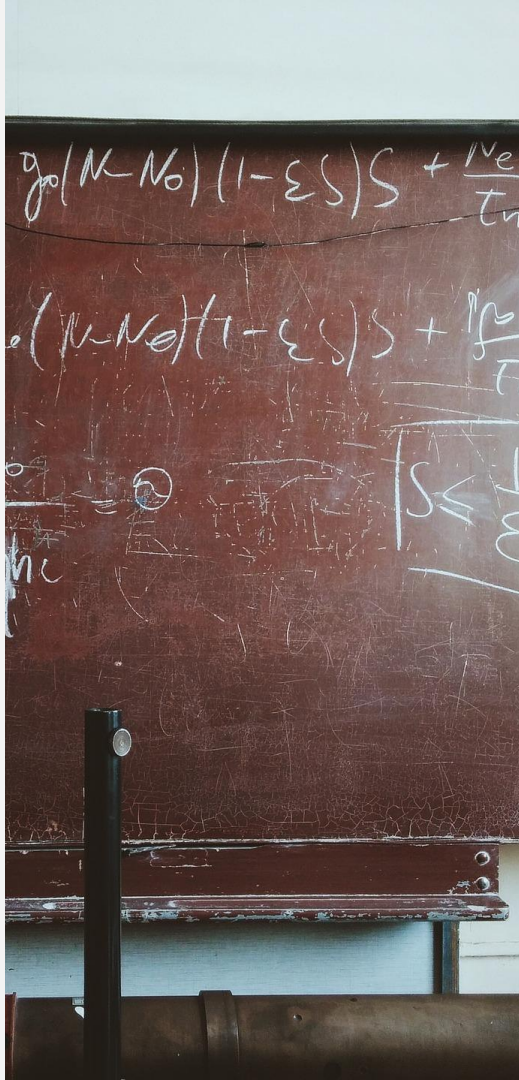


K-fold

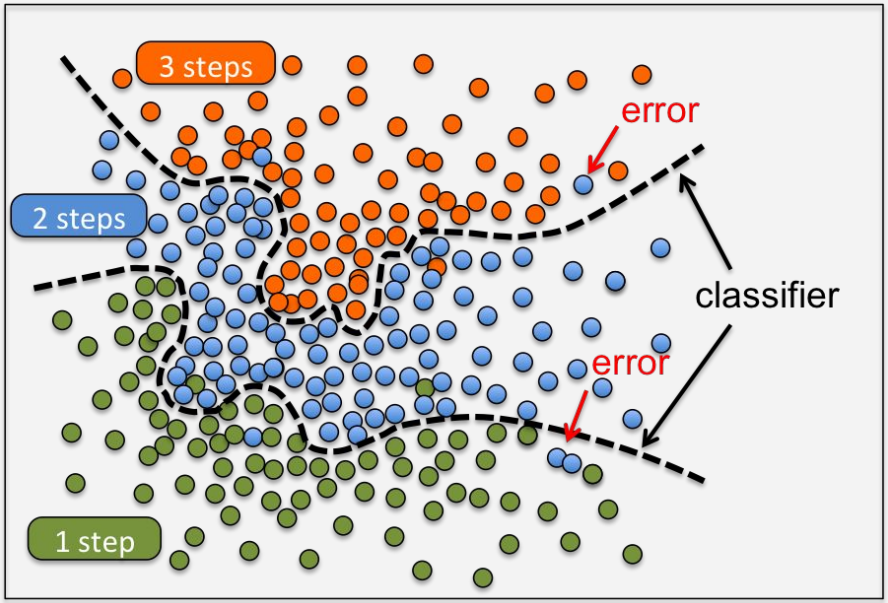

Split 1
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Metric 1
Split 2
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Metric 2
Split 3
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Metric 3
Split 4
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Metric 4
Split 5
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Metric 5
Training data
Test data

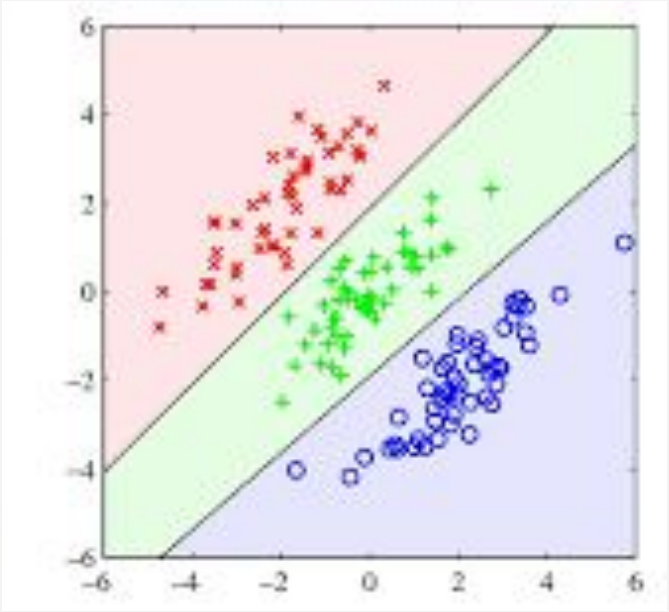
03

Técnicas de Classificação

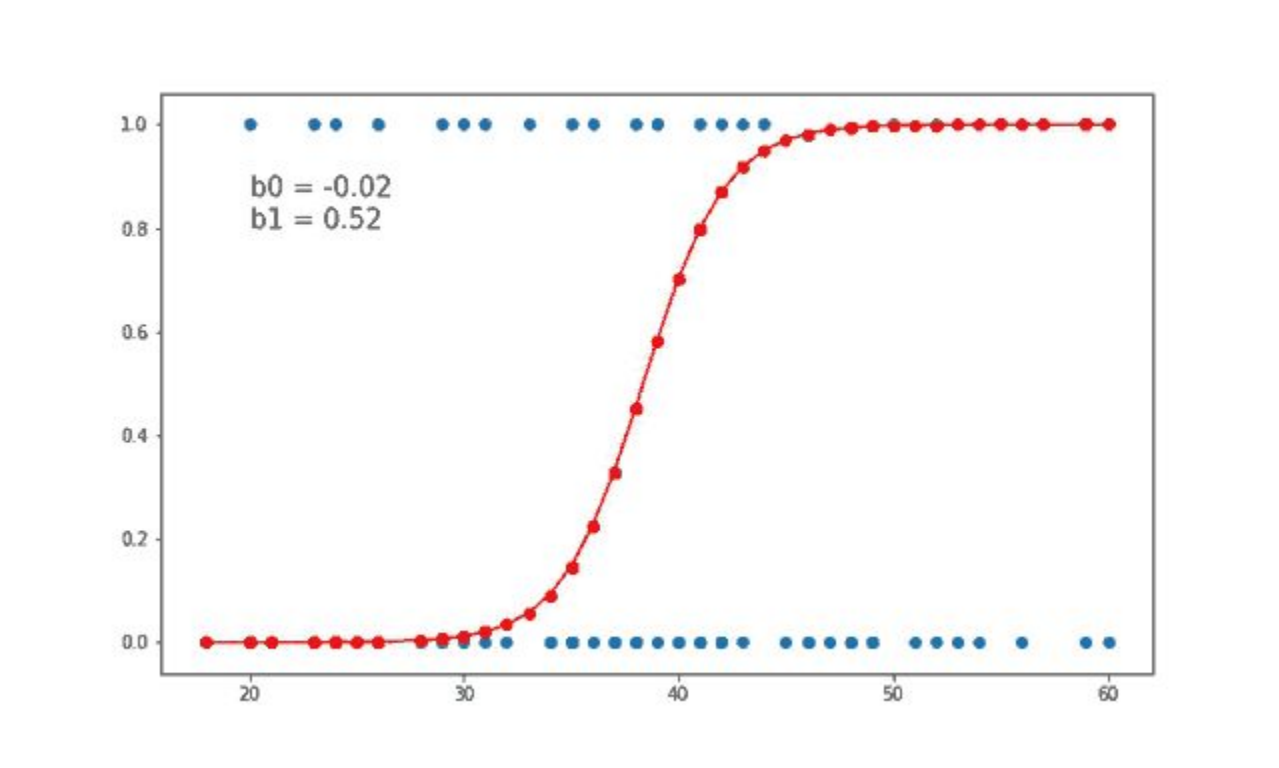


Linearidade

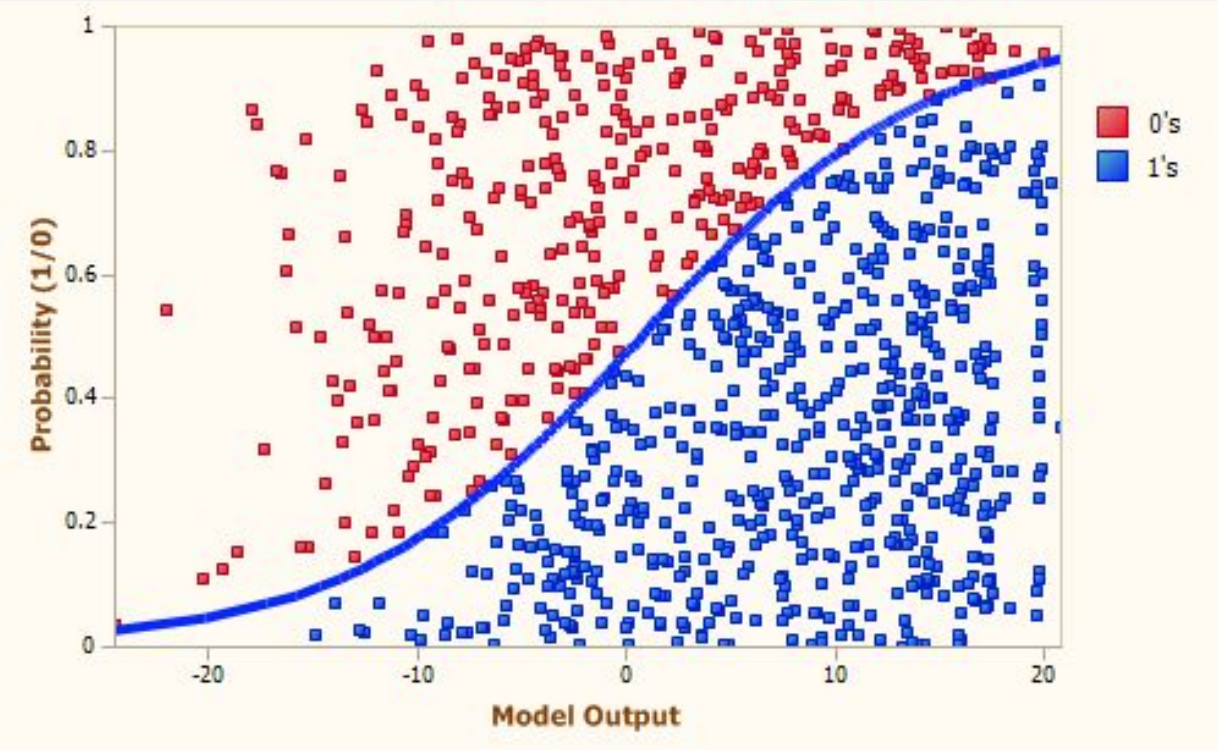




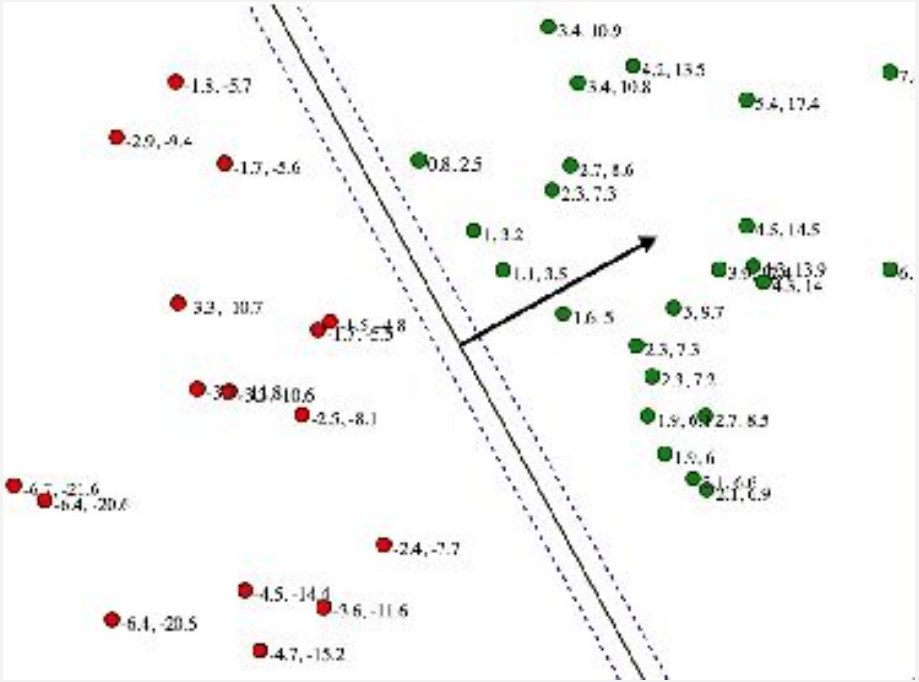
Regressão Logística



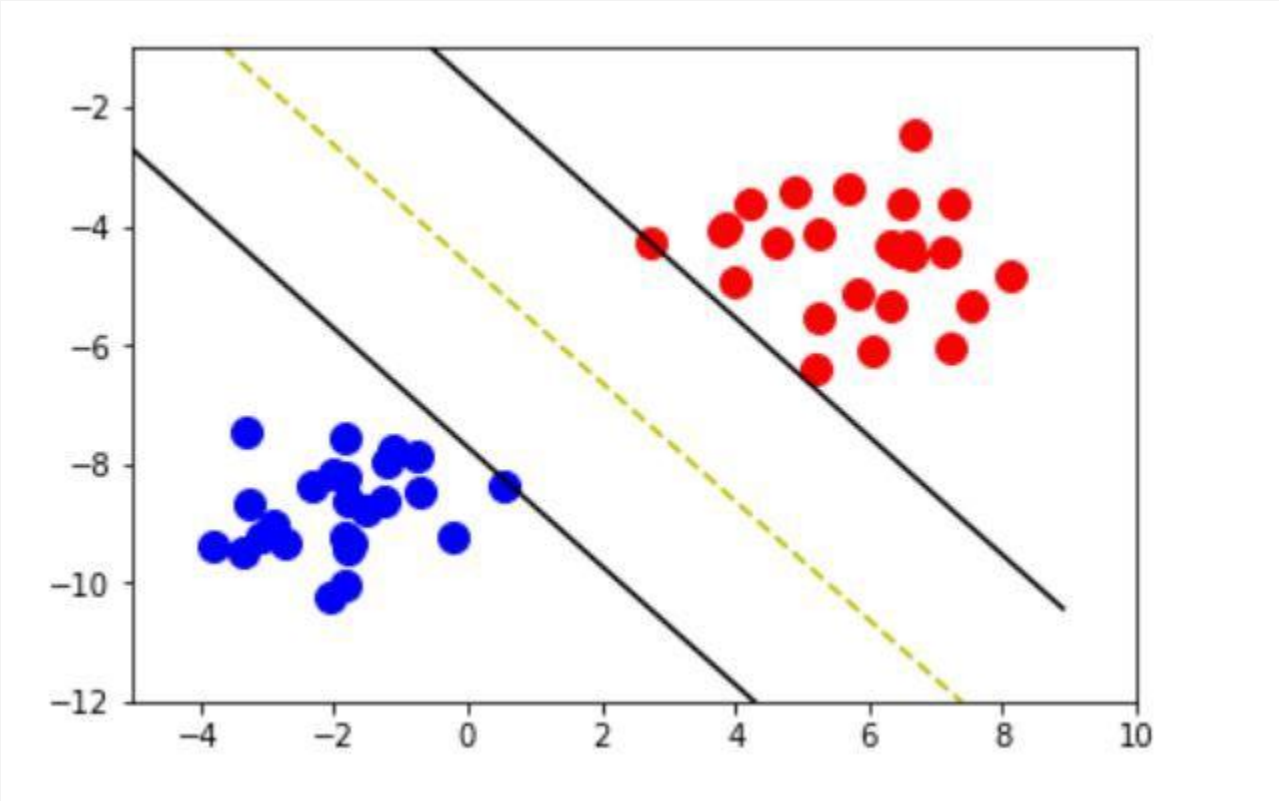
Regressão Logística



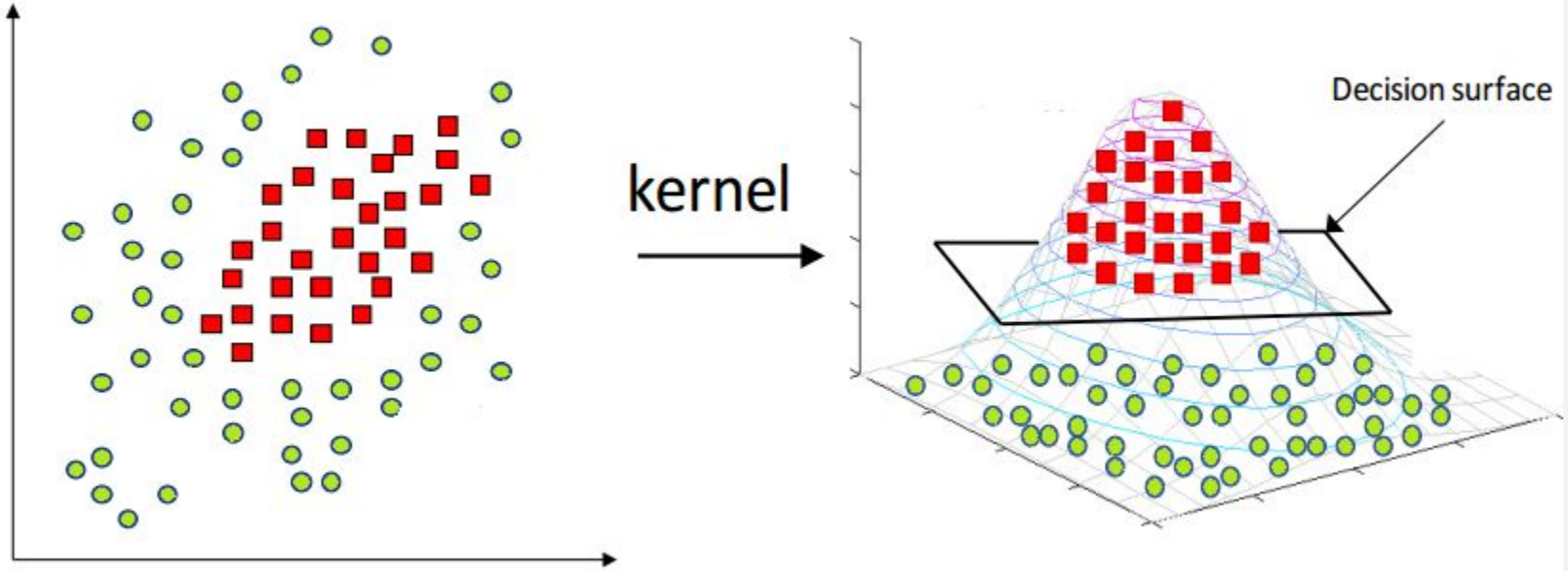
SVM


3.4, 10.9
4.2, 13.5
3.4, 10.8
5.4, 17.4
-1.3, -5.7
-2.9, -9.4
-1.7, -5.6
0.8, 2.5
2.7, 8.6
2.3, 7.3
1, 3.2
4.5, 14.5
1.1, 3.5
13.9
1.6, 5
5, 15.7
2.3, 7.3
2.3, 7.3
-2.5, -8.1
1.9, 6
2.7, 8.5
1.9, 6
2.1, 6.9
-6.7, -21.6
-6.4, -20.6
-2.4, -7.7
-4.5, -14.4
-3.6, -11.6
6.1, -20.5
-4.7, -15.2

SVM - Margens

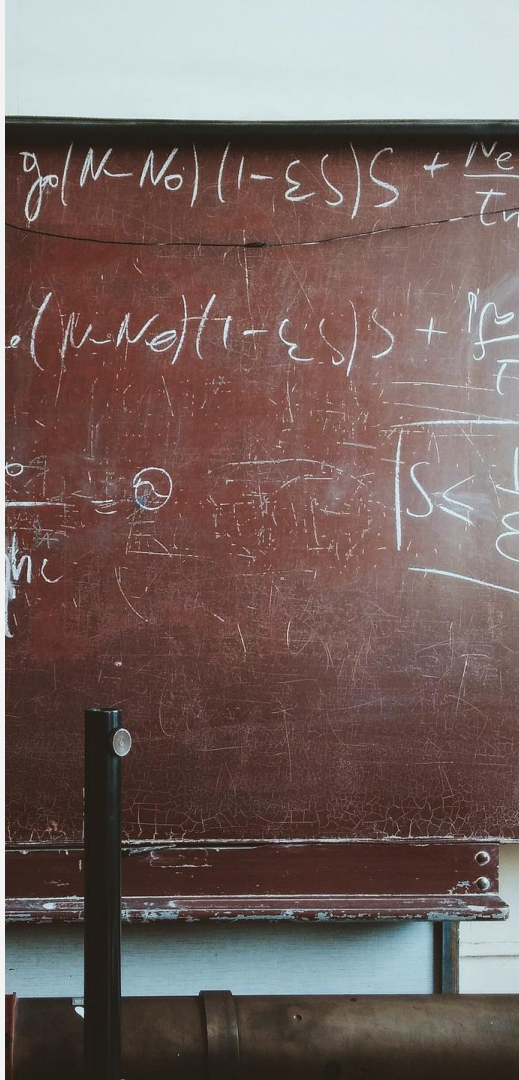


SVM - Truque do Kernel



03

Seu modelo é confiavél?



Matrix da Confusão

		Valor Predito	
		Sim	**Não**
Real	**Sim**	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Acurácia

É a proporção de acertos do modelo

$$Acurácia = \frac{Acertos}{TamanhoDataset}$$

Precisão

É a proporção de verdadeiros positivos?

$$Precisao = \frac{VP}{VP+FP}$$

Recall

Quão precisamente nosso modelo é capaz de identificar os dados relevantes

$$Recall = \frac{VP}{VP+FN}$$